

Segmentation & Profiling Project

Please note that the Data Due Diligence topics/assignments are degree path-specific – i.e., different for MS in Business Analytics and MS in Data Science. Please make sure to select the appropriate version of the assignment.

Business Analytics Students

Keeping the earlier overall description of the Segmentation & Profiling Project, your task is to continue to analyze the Customer Database dataset (used earlier in the Data Due Diligence Project). As before, assume that you are a marketing manager for a telecommunications company and you have been tasked with developing customer segmentation that can support effective, and economically sound, customer retention efforts. To accomplish that goal, you need to take the following data analytical steps:

- Carefully review all available data (i.e., the contents of Customer Database), paying particular attention to distinct data types (e.g., product purchases, demographics, etc.), and identify specific variables that you would like to use as the basis of your segmentation analysis. It is important to be able to justify your choice(s)
- Select a data segmentation technique (e.g., K-Mean Cluster) and segment your customer base
- Evaluate the statistical and practical efficacy of your segmentation solution
- Group all customers into the created segments
- Provide a detailed overview (i.e., the number of segments, cross-segment differences, segment-specific profiles, etc.) of the selected segmentation results, expressly differentiating between 'high' and 'low' value segments; make sure to provide sound justification for why you consider some segments to be 'high value' while others to be 'low value'
- Design and describe explicit customer retention recommendations, based on the results and conclusions arrived at your earlier analytical steps. More specifically, review the value (to the company) of the individual customer segments, singling out those you deem to be most valuable and retention worthy and provide a clear justification of your choices and/or reasoning.

Content-wise, your report needs to include the following elements:

- Executive Summary: 2-3-page summary of your key findings and recommendations
- Segmentation solution selection logic: Description of the rationale you use to select the specific segmentation solution.
- Detailed findings (i.e., detailed customer profiles of individual segments): Descriptions of the individual customer segments.
- Appendix: Any additional details such as tables, charts or figures that you found worthwhile but did not include in other sections.

Format-wise, your report needs to be structured as follows:

- Length: No more than 15 single-spaced pages (excluding any appendices)
- Use Times New Roman, 12-pt font throughout.
- Use clear headings, as appropriate.

- Small blocks of code are acceptable in the main body of text when they improve clarity. Large blocks of code should not be included in the main body of text, include them in the appendices instead.
- Output tables should be cleanly formatted and easy to read. Do not copy and paste directly from console output.
- Format: MS Word.

Data Science Students

Keeping the earlier overall description of the Segmentation & Profiling Project, your task is to continue to analyze the Customer Database dataset (used earlier in the Data Due Diligence Project). As before, assume that you are a marketing analytics manager for a telecommunications company and that you have been tasked with developing customer segmentation that can support effective, and economically sound, customer retention efforts. To accomplish that goal, you need to take the following data analytical steps:

- Carefully review all available data (i.e., the contents of Customer Database), paying particular attention to distinct data types (e.g., product purchases, demographics, etc.), and identify specific variables you would like to use as the basis of your segmentation analysis – make sure to also consider any new variables or indicators you might have previously (Data Due Diligence Project) created, or consider creating new variables now - it is important to be able to justify your choice(s)
- Review the available clustering techniques and select one technique.
- Segment your customer base if you want using k-means clustering approach. If you want to use another technique, please feel free to do so.
- Evaluate the statistical and practical efficacy of your segmentation solution.
- Using the solution, you chose above, group all customers into the created segments.
- Using the created segment, build a classification model to classify these segments that you build during the segmentation solution. You can use Random Forest as shown in class to build the classification model. The dependent model here will be the clusters that you assigned to each observation and will be a nominal variable and you cannot therefore use linear regression, but you can use multinomial logistic regression.
- Provide a detailed overview (i.e., the number of segments, cross-segment differences, segment- specific profiles, etc.) of the selected segmentation results, expressly differentiating between ‘high’ and ‘low’ value segments; make sure to provide sound justification for why you consider some segments to be ‘high value’ while others to be ‘low value’

Content-wise, your report needs to include the following elements:

- Executive Summary: 2-3-page summary of your key findings and recommendations
- Segmentation solution selection logic: Description of the rationale you used to select the specific segmentation solution.
- Detailed findings (i.e., detailed customer profiles of individual segments): Descriptions of the individual customer segments.

- Appendix: Any additional details such as tables, charts or figures that you found worthwhile but did not include in other sections.

Format-wise, your report needs to be structured as follows:

- Length: No more than 15 single-spaced pages (excluding any appendices)
- Use Times New Roman, 12-pt font throughout.
- Use clear headings, as appropriate.
- Small blocks of code are acceptable in the main body of text when they improve clarity. Large blocks of code should not be included in the main body of text, include them in the appendices instead.
- Output tables should be cleanly formatted and easy to read. Do not copy and paste directly from console output.
- Format: MS Word.